ANNUAL REVIEWS

Annual Review of Genetics

Pooled Genome-Scale CRISPR Screens in Single Cells

Daniel Schraivogel,[1] Lars M. Steinmetz,[1,2,3] and Leopold Parts[4]

[1]Genome Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; email: daniel.schraivogel@embl.de

[2]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA; email: lars.steinmetz@stanford.edu

[3]Stanford Genome Technology Center, Stanford University School of Medicine, Palo Alto, California, USA

[4]Wellcome Sanger Institute, Hinxton, United Kingdom; email: lp2@sanger.ac.uk

Annu. Rev. Genet. 2023. 57:223–44

First published as a Review in Advance on August 10, 2023

The *Annual Review of Genetics* is online at genet.annualreviews.org

https://doi.org/10.1146/annurev-genet-072920-013842

Keywords

single cell, CRISPR, functional genomics, genomics, cell sorting

Abstract

Assigning functions to genes and learning how to control their expression are part of the foundation of cell biology and therapeutic development. An efficient and unbiased method to accomplish this is genetic screening, which historically required laborious clone generation and phenotyping and is still limited by scale today. The rapid technological progress on modulating gene function with CRISPR-Cas and measuring it in individual cells has now relaxed the major experimental constraints and enabled pooled screening with complex readouts from single cells. Here, we review the principles and practical considerations for pooled single-cell CRISPR screening. We discuss perturbation strategies, experimental model systems, matching the perturbation to the individual cells, reading out cell phenotypes, and data analysis. Our focus is on single-cell RNA sequencing and cell sorting–based readouts, including image-enabled cell sorting. We expect this transformative approach to fuel biomedical research for the next several decades.

1. INTRODUCTION

If the founders of molecular biology had been asked half a century ago for a dream technology for studying cells, it is not inconceivable that they would have described precisely perturbing thousands of genes in a single experiment and measuring the effects on cellular phenotypes. We have entered an era of genetic experimentation in single cells that is accessible at scale to most laboratories (8). The early applications have yielded basic discoveries about gene function; its variation across signaling states, cell types, and individuals; its role in development and disease; and its interdependence on other genes or molecules. The vast uncharted space of gene function will be the target of study for the next decade. Using these new technologies, we are beginning to assign functions to every single nucleotide in the human genome, across cell types and states, including the noncoding parts of the genome. This endeavor will change our approach to research as well as how we diagnose and treat diseases.

There are two key enablers of these opportunities. First, the CRISPR-Cas system matured into an effective perturbation toolbox. While various genetic perturbation methods have been available for decades, none have the simplicity of use of a target-specifying guide RNA (gRNA) and a constant effector protein, combined with the flexibility of customization of various fusion domains and compatibility with unbiased experimentation at scale. Second, single-cell methods have grown to be highly scalable by miniaturizing and multiplexing the technology. All major omics layers, from chromatin accessibility to messenger RNA (mRNA) and protein concentrations, can be measured, and several have been made accessible to pooled genetic screening (8, 92). Most recently, imaging modalities compatible with pooled measurements have provided a novel view on cell phenotypes (28, 54, 85, 105, 106, 113). With costs dropping by orders of magnitude, typical experiments now provide data on hundreds of thousands to millions of cells.

The final piece of the puzzle is to efficiently link these two powerful advances. The first consideration is that of scale: Combining perturbations of many genes to observe many traits in many cells leads to a small combinatorial explosion of readouts. While pooled single-cell methods are more scalable than arrayed readouts, and their costs have decreased significantly, they are still the limiting factor. The cost of the experiment is one key challenge that we address in this review. The second issue is identifying the perturbation in the readout, a process that we refer to as perturbation genotyping. We focus on scalable pooled approaches, which require a way to identify the perturbation in the assay readout, and outline technical details and potential pitfalls.

This review focuses on recent technologies that probe complex phenotypes from pools of single cells (**Figure 1**). These include single-cell RNA sequencing (scRNA-seq), which captures transcriptome responses along with the perturbation genotype upon pooled perturbation screens (18, 22, 50, 110). We only touch the surface of other single-cell sequencing–based readouts, such as Perturb-ATAC (assay for transposase-accessible chromatin)–seq, or multimodal readouts, such as Perturb-CITE (cellular indexing of transcriptomes and epitopes)–seq (31, 47, 59, 69, 75, 81). The second focus is cell sorting–based readouts, including fluorescence-activated cell sorting (FACS) and image-enabled cell sorting (ICS), that enrich single cells with phenotypes of interest (17, 85). We do not cover readouts without single-cell capture and screens of cell growth, survival, and proliferation (reviewed previously in 8).

2. PERTURBATION TOOLBOX AND EXPERIMENTAL SYSTEMS

We first discuss important practicalities of single-cell pooled genetic screens in mammalian model systems. The key considerations are the choice of the perturbation approach, its delivery, the cell model used, and screen design (**Figure 1**).

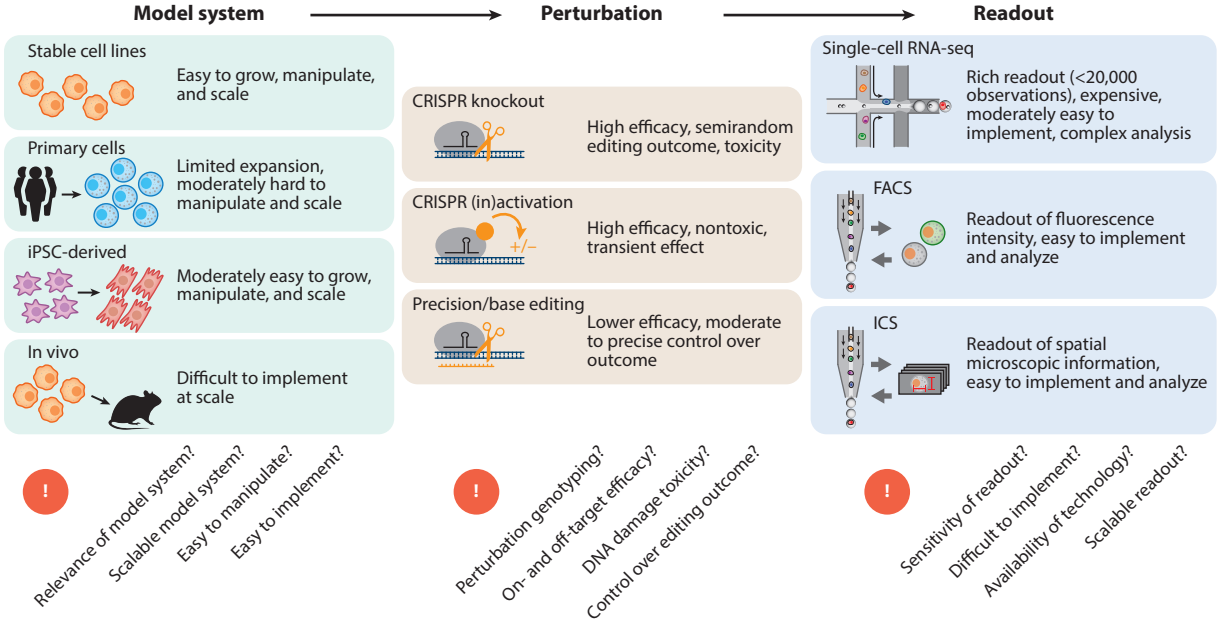

Figure 1

Single-cell pooled genetic screening with complex readouts from single cells; outlines of experimental model systems, pooled perturbation technologies, and single-cell readouts. We focus on technologies discussed in this review, including single-cell transcriptomics and cell sorting–based screens. (*Below*) Outlines of key considerations for such pooled genetic screens. Abbreviations: FACS, fluorescence-activated cell sorting; ICS, image-enabled cell sorting; iPSC, induced pluripotent stem cell; RNA-seq, RNA sequencing.

2.1. CRISPR-Cas Systems

CRISPR systems are scalable and can provide genome-integrated barcodes for perturbation genotyping. They support different perturbation modes (e.g., cutting, nicking, and epigenetic inactivation/activation), CRISPR components (e.g., Cas9, Cas12a, and variants thereof), and guide design principles (120). We outline the key considerations for the decisions below.

2.1.1. CRISPR efficacy. Most pooled screening technologies only read out a sequence barcode associated with the perturbation but not the perturbation state itself (e.g., the editing outcome at the target locus). The key requirement of the perturbation system is high efficacy to ensure that the pooled screen is sensitive in measuring genome function. A main efficacy determinant is a high enough dose of the CRISPR reagents. An easy-to-implement strategy to ensure this is to use Cas expression as a proxy: Cas protein and gRNA expression are correlated, as unbound gRNA and Cas protein are unstable. By labeling the Cas component with a fluorescent tag, cells with low or no Cas expression can be excluded before assay readout. Among the remaining cells with mid- to high-Cas expression, we do not recommend selecting further, since we have not observed correspondingly larger effect sizes as long as substantial amounts of Cas protein and gRNA are present. Using the most recent Cas9 versions helps to increase on-target and decrease off-target effects. The most relevant differences between Cas versions are on- and off-target efficacy, the protospacer-adjacent motif (PAM) sequence, and the length of the expression cassette. Smaller Cas versions, such as SpCas9 and mini-Cas9, are easier to transduce and can fit in a single lentiviral vector with the gRNA (48, 111). Split-Cas9 and mini-Cas9 are compatible with adeno-associated

virus delivery but with the compromise of decreased on-target efficacy (100, 119). Finally, we recommend checking Cas and gRNA expression throughout the experiment. This is particularly relevant during induced pluripotent stem cell (iPSC) differentiation, where most cells can lose Cas expression due to transgene silencing (64).

2.1.2. CRISPR perturbation modes. CRISPR can be used for genome editing as a nuclease (CRISPRn) or to inactivate (CRISPRi) and activate (CRISPRa) transcription (**Figure 1**).

- CRISPRn is efficient and best understood. Its limitations are the semirandom editing outcome and double-strand break–induced stress on the cells. As CRISPRn does not necessarily result in mRNA expression changes, editing efficacy should be tested on the DNA or protein level (89). In combination with DNA-barcoded knock-in templates, CRISPRn can be used for single-cell pooled knock-in screens (80).
- In CRISPRa/i, catalytically inactive Cas recruits epigenetic regulators, such as KRAB, P300, MeCP2, LSD1, VP64, or combinations thereof, to perturb the activity of gene regulatory elements (3, 42, 53, 55, 116). The CRISPRa/i signals can spread along the linear genome, which can lead to false-positive hits if the targeted element (e.g., an enhancer) and response region (e.g., a promoter) are in close proximity (33, 77, 84). The ability to detect essential genes is similar between CRISPRi (targeting promoters) and CRISPRn (targeting the translated region) (82). One advantage of CRISPRa/i over CRISPRn is that CRISPRa/i changes gene expression, allowing a readout of the perturbation effect on top of the perturbation genotype via scRNA-seq (77).
- Specific sequence changes can be installed using CRISPR systems that increase the likelihood of homology-dependent repair, base editors with dCas9, or prime editors with gRNA-encoded changes (5, 12, 29). This allows the functional characterization of defined genetic variants or saturation genome editing, for example (26). Base editors can achieve almost 100% editing rates in cultured cells and have been combined with FACS-based screening (14, 46). Since a single base editor/gRNA combination can edit multiple nucleotides, computational approaches are needed to deconvolute gRNA-level to nucleotide-level measurements (14). Prime editors, while precise and already used in pooled screens, remain less efficient for now (26, 57).

Additional CRISPR systems for pooled single-cell screening process their own RNA component (crRNA) out of larger precursors, termed CRISPR arrays. These CRISPR arrays encode multiple crRNAs as a single polymerase III transcription unit to target the same gene with multiple crRNAs for increased on-target efficacy or to target different genes for combinatorial screens. These systems include Cas13d and Cas12a, which have been combined with scRNA-seq and FACS for combinatorial screening (36, 107).

2.1.3. Guide RNA library design. The gRNA libraries should be designed using the latest design algorithms and contain positive and negative controls. For CRISPRn/a/i of transcribed coding and noncoding genes, gRNA design principles are mature (24, 43, 82). For CRISPRa/i of regulatory elements other than promoters, it is generally advisable to place gRNAs in open chromatin regions, such as DNase-seq or ATAC-seq peaks; other than that, there is no favorite gRNA design tool that outperforms others. For enhancers, gRNAs seem to either function or not (33, 84), and it makes sense to target these elements with more guides than usual. We recommend validating gRNA designs using some example gRNAs in a given cell type and, as a last resort, treating the gRNAs as technical replicates, using enough ($n \geq 3$) gRNAs per target element to allow the filtering of nonfunctional gRNAs by lack of an on-target effect. Off-target effects can be checked by assessing changes at nearby genes of the off-target loci (74).

Most frequently, a single gRNA is tested per cell. This minimizes stress from CRISPR-Cas activity and prevents potential dilution effects due to limited Cas protein availability. However, it can be advantageous to apply multiple perturbations per cell (37, 72, 76, 77, 115). First, it is cheaper. The same cell can be used to test more than one perturbation, as the risk of overlapping effects in the same cell is low in genome-scale screens. This approach is more advisable for CRISPRa/i, as no double-strand break stress is generated. Second, combinatorial perturbations are generated on purpose to detect epistatic effects or genetic redundancy. To this end, gRNA combinations can be applied in a randomized fashion or by generating defined sets of perturbations (11, 20). Cloning strategies to generate such multitargeting libraries have been reviewed previously (66).

2.2. CRISPR-Cas Construct Design and Delivery

While perturbation efficacy is crucial for overall screen quality, it is also central to the design and delivery of the CRISPR constructs. Monoclonal cell lines robustly express Cas protein from random integration with lentivirus at low copy and transposases at high copy or produce controlled single-copy integration into a landing pad (6, 57). These strategies are laborious but form a clean foundation for pooled screens. For speed, cost, and convenience, polyclonal cell lines with random integrations in every cell can provide sufficient-quality data (103). gRNAs are most often delivered in a viral vector into a Cas-positive cell line or into a naive one followed by transfection with a Cas plasmid, mRNA, or protein to deliver a transient high dose of the enzyme. The latter is only compatible with CRISPRn or a short time course on transcriptional modulation. The efficacy of delivery is an important consideration, as cell lines vary broadly in their capacity to be infected or transfected (112). We recommend accurately quantifying the efficacy of delivery and survival after perturbations to ensure realistic coverage estimates and good control over the number of viral integrations per cell.

Another important consideration is the choice of the selection method: Fluorescent reporters allow fast selection with FACS, while antibiotics vary in selection strength. For gRNA delivery, puromycin is powerful since it allows fast and strong selection of complex gRNA libraries. We further recommend directly selecting for Cas expression rather than using selection markers expressed from a different promoter. This ensures optimal control over the Cas dose and the ability to exclude transgene silencing. In many systems, such as iPSCs or primary T cells, the perturbation itself (e.g., the DNA break) is negatively selected against, reducing coverage. This is especially true for antibiotic selections where a few cells that survived selection without stably expressing the transgene will quickly overtake the population. We recommend thoroughly understanding the makeup of the postselection cell pool and losses at each stage of the delivery and selection process.

Single-cell pooled screens are always growth screens since some perturbations will have negative effects on cell growth, resulting in lower abundance or loss of these perturbations from the readout population. Coverage on fitness genes can be increased with inducible expression systems or by sampling early time points before the perturbation drops out. Alternatively, sensitivity for these genes can be improved by overrepresenting their gRNAs in the library (77).

2.3. Cell Model Systems for Single-Cell Screening

Single-cell readouts and perturbations at scale are compatible with most cell model systems. The major restrictions are compatibility with the selected readout, susceptibility to the perturbation reagents, and complexity of the model.

First, the readouts of scRNA-seq and cell sorting set limitations that depend on morphological properties of cells. For example, large cells, such as cardiomyocytes, might not be compatible with

droplet- and nanowell-based scRNA-seq. In such cases, working with extracted nuclei or differentiated but not terminally matured cell stages can increase compatibility. In addition, some cell types show low recovery after droplet-based scRNA-seq for yet unknown reasons (83). FACS/ICS readouts are similarly limited by the nozzle size, which is the smallest restriction in a cell sorter that restricts assays to particles ~100 μm in diameter at reasonable sorting speed. And cells can be sensitive to the shearing forces and pressure changes during FACS-based sorting, which can lower the yield of intact cells.

Second, to achieve sufficient coverage of the pooled gRNA library, often millions of cells need to be transduced. Standard stable cell lines generally transduce and cultivate efficiently, and easily scale with the throughput needed for pooled screening. Primary cells are more difficult to transduce, especially with large Cas expression constructs. In addition, primary cells often have limited cultivation time, and consequently no validated stable monoclonal cell lines can be established. In these cases, polyclonal cell lines have proven useful, although with CRISPR efficacy restraints. Much progress has been made in primary T cells, for example (83, 87, 88, 98). Another frequently used model system is iPSC-derived cells, and pooled single-cell screening with scRNA-seq and FACS readout has been established for multiple differentiation protocols (25, 97). For iPSCs, the following strategies have proven useful: (*a*) applying Cas and gRNAs on the iPSC stage before initiating differentiation, (*b*) preventing transgene silencing through targeted integration of Cas into a safe harbor locus, and (*c*) validating monoclonal cells for Cas expression and CRISPR efficacy (25, 97).

Finally, in cell models where multiple cell types and states are present in variable ratios, such as primary cell material and organoid models, sampling enough cells to recover effects across cell states can be difficult. These multiple cell states constitute de facto subclones, each of which has lower coverage. To this end, it is recommended to (*a*) increase cell coverage per gRNA, (*b*) distinguish cell states via surface markers (for FACS readouts) or RNA expression (for scRNA-seq readouts), or (*c*) use clonal barcoding to label individual lentiviral integration events of the gRNA expression construct (68).

2.4. Designing and Optimizing the Screen

Given all the considerations above, how should one set out to implement a pooled single-cell CRISPR screening campaign?

2.4.1. Start step by step. We recommend running small-scale pilots consisting of 20 to 50 positive and negative controls. These positive controls are perturbations with known effects/targets, as well as effect sizes. These pilots aim to answer several questions: Can I detect the perturbation? Do I achieve strong on-target effects? Can I scale my model system? What is the best time point to read out perturbation effects? How many gRNAs, cells, and reads do I need per target element? It is advisable to select positive controls from a range of different effect sizes (weak versus strong) and target gene expression levels (low versus high). Weak effects of lowly expressed genes are more difficult to detect. We recommend high cell and sequencing coverage for these pilots, allowing downsampling analyses to determine optimal cell numbers and sequencing coverage (84). We recommend at least 100 cells per gRNA, and at least 30,000 reads per cell, and then using these pilot data to make informed coverage decisions. As negative controls, we recommend both nontargeting and targeting gRNAs. For CRISPRn, multitargeting gRNAs can estimate the impact of excess DNA breaks on assay performance. If no positive controls are available, gene regulatory networks or publicly available data sets [such as the Cancer Dependency Map (DepMap) (7, 101)] can help prioritize genes. The ideal time point of the readout is not known a priori, and we recommend

using the pilot screen, or, alternatively, an arrayed pilot, to evaluate this. A rough estimation for CRISPRa/i effects is 4 to 5 days to reach a maximum effect (64) and 1 to 2 days for CRISPRn (94).

2.4.2. Scaling up. For large-scale screening, the controls from the high-coverage pilot should be included across replicates and batches, allowing quality checks with known perturbation effects. Overall, the more positive controls there are, the better. Throughout the screen, from gRNA library generation to cell growth and readout, bottlenecks should be avoided; otherwise they result in excessive loss and unequal gRNA representation. The technical quality of the screen is determined by the strongest bottlenecks. The first bottleneck occurs during library delivery, with >200× coverage per gRNA recommended, but as low as 50× coverage per replicate is acceptable (23). Each step, from passaging to DNA extraction and library preparation, should have ample coverage (>500×). For guidance on estimating ideal coverage from DNA library quality, we recommend Reference 23. The second bottleneck occurs during the readout. Generally, one wants to process only as many cells as necessary, but enough cells to achieve sensitivity. Sensitivity in single-cell screening largely depends on the number of cells retrieved per perturbation for assay readout but also on the complexity of the system, where, for example, the cell cycle stage can create internal grouping of cells and result in lower coverage. In scRNA-seq screening, several hundred thousands of cells can easily be processed in a single step, excluding cell handling as a limiting factor. However, single-cell library preparation and sequencing are expensive, and it makes sense to think about the necessary sensitivity of the screen, guided by a pilot experiment. We describe in Section 3.1 how to reduce the necessary cell and sequencing coverage, for example, using targeted readouts. The consumable costs of running cell-sorting screens are low. However, depending on the event rate (i.e., the number of cells that can be phenotyped and isolated per second), target population frequency (i.e., the percentage of events that are of interest), and library size, sorting can take many hours of pure instrument run time.

2.4.3. Scaling down. Given the scale and complexity of pooled single-cell screens, it is appealing to reduce their size by measuring fewer perturbations or readouts and computationally imputing the rest. This approach is supported by the structure evident in the screen results, where very similar effects can be observed for multiple perturbations or different genes respond in synchrony across all of them. The concept of compressing perturbations has been used from model organism genetics (21) to cancer cell line characterization and relies on statistical reconstruction to balance experiment size against the chosen objective, for example, to explain variance in all outcomes, measure representatives in many categories, or even simply randomly choose the genes (15, 104). Analogously, the outputs can be focused to obtain high-quality data from a key set of genes or to measure only an informative subset that allows the rest to be predicted (84, 93). In each case, however, measurements are substituted for statistical assumptions about the data, and the results will suffer if these assumptions do not hold.

3. POOLED CRISPR SCREENS WITH scRNA-SEQ READOUTS

In this section, we describe pooled screening readouts using scRNA-seq. We compare scRNA-seq platforms, outline how to link perturbations with transcriptomes, and describe computational strategies to read out perturbation effects from scRNA-seq data (**Figure 2**).

3.1. Single-Cell Transcriptomic Platforms for Pooled Screening

The core protocol for pooled scRNA-seq screening is droplet-based single-cell sequencing (10, 49, 56, 62, 79, 121). The transcriptome is captured from cells or nuclei and read out from either

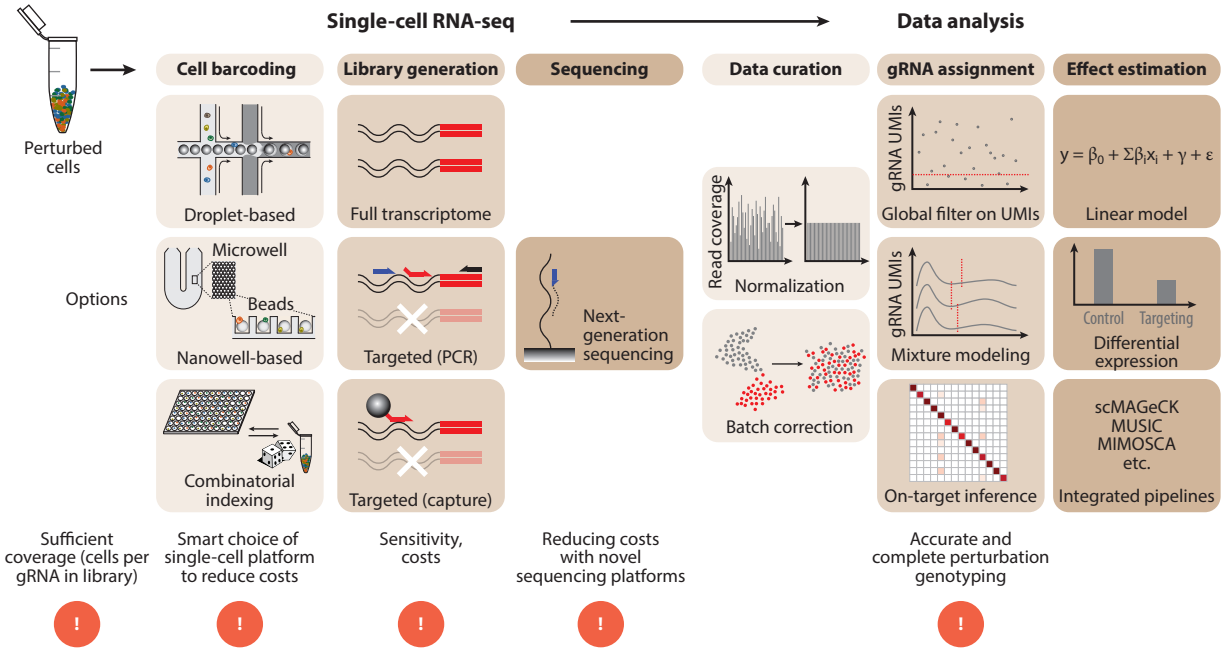

Figure 2

The workflow of CRISPR-Cas9 screens with single-cell RNA-seq readouts. Cells that were perturbed in a pooled fashion with CRISPR-Cas9 are the input material for single-cell RNA-seq technologies. Major options are the single-cell technology (cell barcoding) and library generation strategy. Libraries are sequenced with Illumina or novel sequencing technologies that reduce sequencing costs. Data are analyzed in several steps, including data curation, gRNA assignment, and differential gene expression testing. Options exist for perturbation genotyping and differential gene expression testing. Abbreviations: gRNA, guide RNA; PCR, polymerase chain reaction; RNA-seq, RNA sequencing; UMI, unique molecular identifier.

the 3′ or 5′ end. Either the full or targeted transcriptome consisting of several hundred RNAs of interest is read out. The basic approach has been reviewed elsewhere (13, 95).

Current developments focus on improving the single-cell readout in several ways. For example, targeted perturbation sequencing (TAP-seq) increases sensitivity and reduces cost by selectively amplifying up to 1,000 selected transcripts per cell (84) (**Figure 2**). This strategy improves signals from those genes at the expense of the rest of the transcriptome. Alternatives to this amplification-based method are oligonucleotide capture approaches (76) or hybridization-based techniques (65). Another way to improve the signal from cells is to use gRNA unique molecular identifiers (UMIs) that support tracing perturbation effects in many parallel cell pool subpopulations for replication. Cost is further reduced by multiplexing multiple gRNAs into each cell, assuming each to rarely have large effects and thus compressing their effect measurements into one to assay fewer cells without a signal. The price of single-cell library preparation is further reduced by innovation in cell barcoding methods, including nanowell-based systems (e.g., BD Biosciences Rhapsody HT Xpress), combinatorial indexing (79, 90), and commercial derivations thereof (e.g., Parse Biosciences (99) and Scale Biosciences). Much-awaited reinvigorated competition promises the same for short-read sequencing (e.g., provisions by Ultima Genomics or Element Biosciences).

Another consideration for the choice of the scRNA-seq platform is the ability to prepare multiple sequencing libraries from the same cells: Several technologies capture the transcriptome on solid beads (e.g., Drop-Seq, BD Biosciences Rhapsody), allowing multiple libraries from the same copy DNA (cDNA) to be prepared since the original unamplified cDNA is stably attached to a

bead and can be recycled. This is different from 10x Genomics and combinatorial indexing strategies, where the user has a single shot to amplify the transcriptome for library preparation. This, for example, allows a whole transcriptome library to be prepared to observe global gene expression effects, followed by a targeted transcriptome library to increase sensitivity to genes of interest.

3.2. Perturbation Genotyping

Most scRNA-seq methods are selective for polyadenylated RNA and sequence the 3′ and 5′ ends of mRNAs. However, gRNAs are expressed as polymerase III transcripts. Consequently, gRNA transcripts do not contain a poly-A tail and cannot be captured using most scRNA-seq methods. Three strategies have been established to solve this problem.

3.2.1. CROP-seq. CRISPR droplet sequencing (CROP-seq), a frequently used strategy for perturbation genotyping with scRNA-seq, is based on a specific lentiviral vector design (18, 73). The gRNA expression cassette, consisting of a polymerase III promoter, gRNA, and a terminator, is located within the 3′ long terminal repeat (LTR) of the CROP-seq vector. A second transcription unit initiates from a polymerase II promoter and expresses the selection marker. Since this polymerase II transcript terminates in the 3′ LTR, the gRNA cassette becomes part of the 3′ untranslated region (UTR) of the lentiviral polymerase II transcript. Consequently, 3′ scRNA-seq can capture the gRNA locus, and the gRNA itself serves as a barcode for perturbation genotyping. This elegant system prevents the need for a dedicated barcode that is associated with the gRNA, and CRISPR efficacy using an optimized CROP-seq vector is as high as with traditional gRNA expression constructs (84). There are two disadvantages of CROP-seq: First, the length of the polymerase III cassette that can be integrated into the 3′ LTR without affecting its function is limited. Suboptimal 3′ LTRs result in low lentiviral titers and inefficient proviral integration. Therefore, adding multiple gRNA expression cassettes for combinatorial editing might exceed this limit for 3′ LTR insert length. Second, perturbation genotyping is not successful for all cells; that is, the gRNA is detected in a subset of cells only. For K562 cells, this subset is around 30% and seems to depend on the cell type and scRNA-seq platform. Targeted amplification of the gRNA-containing polymerase II transcript solves this issue, allowing retrieval of the gRNA identity from over 90% of cells when using CROP-seq (1, 41, 84).

3.2.2. Dedicated sequencing barcodes. Three other methods, Perturb-seq, CRISP-seq, and MOSAIC-seq, make use of dedicated barcodes for perturbation genotyping (1, 50, 110). These barcodes are short sequence stretches that are located close to the 3′ end of the lentiviral polymerase II transcript but separate from the polymerase III gRNA locus. After the dedicated barcode and gRNA have been cloned into lentiviral backbones, barcodes and gRNAs need to be linked by next-generation sequencing. Since the dedicated barcode is close to the 3′ end of the polymerase III transcript, 3′ scRNA-seq allows perturbation genotyping. There are two major disadvantages of such dedicated barcodes: First, during gRNA library cloning, lentiviral production, and transduction, gRNAs and barcodes can shuffle. The rate of shuffling increases with the distance between the gRNA locus and barcode locus. This gRNA–barcode uncoupling can affect as many as 50% of all cells, resulting in false perturbation genotyping and a decreased ability to identify differentially expressed genes (41). Since the major source of this shuffling seems to be the lentiviral reverse transcriptase (which often switches between templates and thereby generates chimeric vector genomes), this effect can be diminished by using a modified lentiviral packaging protocol with a carrier plasmid that does not contain a gRNA (2, 27). Although almost completely abolishing gRNA–barcode shuffling, this strategy results in decreased lentiviral titers. Second, similar to CROP-seq, Perturb-seq and related methods depend on polymerase chain

reaction (PCR) amplification of the barcode transcript to achieve a high proportion (around 90%) of genotyped cells.

3.2.3. Dedicated RNA capture sequences. A third strategy adds the ability of capturing nonpolyadenylated transcripts to scRNA-seq. Custom oligonucleotide sequences were added in addition to oligo-dT sequences to capture beads of commercial scRNA-seq platforms, such as 10x Genomics and BD Biosciences Rhapsody. These capture sequences bind polymerase III gRNA transcripts directly, using either a modified tracrRNA that contains a sequence complementary to the bead capture sequence (3′ capture) or a reverse transcription oligonucleotide complementary to the constant tracrRNA sequence (5′ capture) (76). The major limitations of direct 3′ RNA capture are twofold: First, tracrRNA modifications need to be tested and optimized to achieve high CRISPR efficacy. Second, libraries need to be generated with the modified tracrRNA sequence and will not be compatible across single-cell platforms.

3.3. Principles of Data Analysis

Data generation at scale requires analysis methods that scale accordingly and ideally also provide rigorous guarantees. Guidelines on the analysis of scRNA-seq experiments and pooled genetic screens have been thoroughly reviewed elsewhere (8, 92), and we highlight the main considerations here (**Figure 2**). The primary goal is to accurately estimate differential gene expression effects across cells with different gRNAs. For this estimate to be useful, it has to be comparable to other cells and different experiments. Normalization and correcting for covariates (including batch) are the steps that ensure this property, and the flexibility in available options has to be handled judiciously. We recommend using nearly excessive plotting for comparisons at each stage of raw data processing to ensure that no steps generate additional biases and to evaluate the quality of the preprocessing approach by assessing the performance on preselected positive and negative controls, as well as to make sure that technical and biological replicates are concordant in the end.

The analytical step most specific to pooled scRNA-seq screening is assigning perturbations to individual cells. There are plenty of subtleties, however. For example, if reads that map to multiple gRNAs are found in a cell at different copy numbers, how should we evaluate which perturbation was active? Is a gRNA label enough to say that the perturbation happened, or do we need to test whether the target was successfully perturbed? If we do not observe the gRNA but we do observe a perturbation at a target consistent with the gRNA, is this enough to assign the perturbation state to the cell as if the gRNA had been present? How should one handle partial perturbations, for example, incomplete knockdown in CRISPRi? We offer some guidance below, but the answers are often study specific.

After the data are cleaned and gRNAs are assigned, perturbation effects are measured. The total average effect can be calculated by combining the phenotypes across cells with the same gRNA and contrasting them against controls. The correct model for gene perturbation effects may vary. We recommend checking the measurements for remaining heterogeneity due to covariates (cell cycle, sample metadata, and technical effects), as well as on-target efficacy (Cas expression, variable gRNA efficacy, and per-cell perturbation strength). A natural way to achieve this is in a linear modeling framework with regularizing priors, using principles developed for differential expression analysis as a practical guide (61). Many visualizations of each perturbation-target pair are needed to confirm that no confounding signal remains. A general approach is to plot the positive and negative control distributions of the target expression or other chosen effect statistic across cells and to make one panel for each of the recorded cell covariates (e.g., batch, cell cycle stage, and gRNA used), stratifying the cells with the perturbation by that covariate.

3.4. Data Analysis Step by Step

The push for reproducibility and open science has encouraged the useful sharing of analysis workflows, and entire books have been compiled that aid single-cell data analysis (40).

3.4.1. Filtering, normalization, and covariate correction. There are many options available for common data processing tasks (4, 118). For example, one popular choice is using 10x Genomics Cell Ranger to provide initial steps, while Seurat, Scanpy, or scvi-tools can be used for quality control to derive a useful data matrix and for fitting linear models of effect or dimensionality-reduction models to explore the structure in the data (34, 38, 108). As a reasonable starting point, we recommend filtering cells based on total UMI count and mitochondrial content, normalizing to 10,000 UMIs per cell, and correcting for cell cycle and other known experimental covariates. It is also useful to monitor low-dimensional embeddings of the data (e.g., UMAP, t-SNE, or PCA) with annotated quality-control metrics and covariate states to confirm that confounding sources of signals are appropriately removed.

3.4.2. Perturbation genotyping. We assume that a count matrix of gRNA reads in each cell is given, a review is performed, and these data are then used for gRNA assignment. One broad class of approaches uses the absolute number of reads in one cell to call gRNAs by filtering them based on thresholds for minimum UMI occurrence. This natural method is implemented by default, for example, in Cell Ranger, and variations on this theme have been broadly used (76, 84). However, it can be difficult to get the thresholds right: Coverages and gRNA expression efficacies vary between experiments, and ambient gRNA transcripts can give rise to ectopic UMIs at different rates, rendering the thresholds incompatible between cells and not transferable between studies. Thresholding can also easily assign many gRNAs to one cell, which can be unrealistic or undesired, and require postprocessing, for example, using the methods described in the next section.

A general alternative approach is to use a model across many cells to identify either the two expected modes of gRNA presence or a perturbed target state. A gRNA can be present at ambient levels with a small number of reads with an expectation of 0 in most cells or expressed at a high level with a large number of reads. Fitting a two-component mixture model and assigning cells to the components result in gRNA calls that are informed by the global expression characteristics of the guide in the experiment (76). As an alternative, the perturbation target expression can look similar to negative control cells, or show a perturbed state reflected by differing sequence or expression. While using the target to identify the perturbation can improve perturbation genotyping, it has only worked robustly for the strongest of effects. We recommend the mixture modeling of gRNA UMI counts as a practical and useful starting point for perturbation assignment.

3.4.3. Perturbation effect estimation and downstream analyses. The final step after gRNA assignment is perturbation effect estimation (**Figure 2**). This is treated as a statistical problem, where the repeated observations of the effect in many cells are contrasted against controls to derive the effect value, while accounting for appropriate confounders. This step is supported in Seurat, scMAGeCK, and Scanpy, while many studies opt for the flexibility of a generic linear mixed-modeling framework or its adjustments (31, 73, 77, 114). All standard differential signal analyses, from simple t-tests on individual genes to gene set enrichment analyses, are sensible options to further explore to understand the impact of the chosen perturbations. As with each step of the analysis, dimensionality reduction methods should be used to check the impact of confounders. At this step, they can also be useful exploratory tools to rapidly evaluate potential causes of structure in the data, but we recommend testing the derived hypotheses directly (e.g., mitochondrial gene perturbations have similar effects distinct from others) rather than deriving clusters and characterizing their properties.

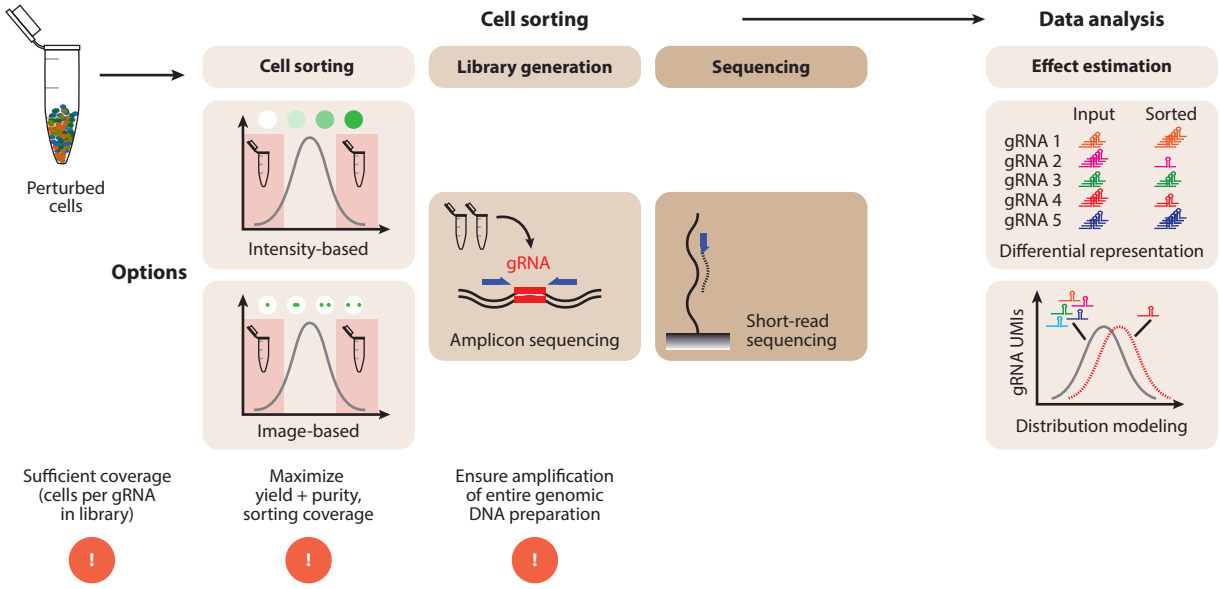

Figure 3

Workflow of FACS- and ICS-based pooled genetic screens. Perturbed cells are used for cell sorting with FACS or ICS. The phenotypic parameter, which is either an intensity-based (FACS) or imaging-based (ICS) parameter, is binned and populations are sorted until sufficient coverage of the gRNA library is achieved. Targeted sequencing of genome-integrated barcodes generates count matrices for each sample for hit calling using classical growth-based CRISPR screening algorithms or specialized tools such as distribution modeling with MAUDE. Abbreviations: FACS, fluorescence-activated cell sorting; gRNA, guide RNA; ICS, image-enabled cell sorting; UMI, unique molecular identifier.

3.4.4. Separation of causes from correlations. Pooled perturbation screens give a window, which is not accessible from observational data, into the causality of gene effects and its percolation to the entire transcriptome. Two types of causal effects are reasonably straightforward to measure: The impact of the perturbation on the target is the change in expression of the perturbed gene, while the total causal effect is the average change in the transcriptome as a result of the perturbation. However, the direct impact of the perturbed gene on other downstream traits, which is perhaps the most interesting effect, requires disentangling from the rest of the changes, as the causal path could also go via multiple intermediates. Inferring the structure of the regulatory network and the strengths of individual links remains an active area of study (35, 52, 60).

4. POOLED CRISPR SCREENS WITH CELL SORTING

We next focus on pooled CRISPR screens with cell sorting enrichment. In particular, we discuss FACS and ICS, which measure protein levels and protein localization, respectively (17, 85, 86) (**Figure 3**). While FACS-based CRISPR screens can identify genes or genetic loci that regulate the expression level of a protein of interest, ICS-based screens identify those that regulate protein localization.

4.1. High-Throughput Cell-Sorting Platforms

FACS extracts up to 40 parameters from each cell, including label-free quantification using scattered light and the intensity of fluorescently labeled proteins, DNA, or RNA. The signal measurements, including the signal area, height, and width, are highly quantitative across multiple

orders of magnitude (17). For genetic screening, only a few fluorescent channels are generally needed (see Section 4.2). The molecule of interest is made detectable by binding to a fluorescent dye or antibody, endogenous tagging, or ectopic expression as fluorescent protein fusion. Cell sorters are broadly available and allow sorting speeds of over 10,000 cells per second.

While traditional FACS is blind to spatial changes in protein localization, ICS expands the parameter space of traditional FACS to spatial measurements by coupling FACS with imaging and image quantification (85). Image reconstruction and quantification happen in real time on the machine, allowing cell sorting according to image-derived parameters at the same speed at which traditional FACS operates. Since ICS uses traditional FACS technology, both technologies have the same sample requirements. In addition to ICS, pooled genetic screens with imaging readouts were enabled by optical tagging of cells followed by FACS (39, 54, 113), in situ sequencing under a microscope (28, 32, 106), or deconvolving surface epitope combinations (109) (reviewed in 105).

4.2. Experimental Design of Sorting-Based Screens

Key technical considerations for the sorting phase of FACS- or ICS-based screens are the sorting speed, stability of the run, stability of the phenotype, purity/yield of sorting, and cell recovery after sorting.

4.2.1. Establishing an optimal sorting strategy. For cell-sorting screens, a classification strategy (also called gating or sorting strategy) needs to be established to isolate cells with high purity and yield. High purity improves the sensitivity of the assay to detect perturbation effects, while high yield decreases sorting time. Sorting time can be calculated from the event rate, gRNA library size, and target population frequency (85). Different cells behave differently in a fluidic stream, and large cells result in lower sorting rates since fewer cells can pass the machine at the same time.

Finding the ideal sorting strategy is more or less straightforward with FACS: In most cases, single cells are first distinguished from debris and cell multiplets using light scatter parameters. Then, cells are distinguished according to the actual phenotype of interest using the area signal (i.e., the overall amount of fluorescence from a single cell) of the labeled protein. Optional filters use viability stains to exclude dead cells and cell cycle stains to focus on a single cell cycle phase.

With ICS, image-derived parameters are used to enrich spatial phenotypes of interest. Depending on target protein expression variability, it can make sense to minimize confounding effects of signal intensity by focusing on cells from a narrow window of the traditional signal area as input for the image analysis step. Finding the ideal parameters for a phenotype of interest can be challenging with ICS owing to the many available options, especially for complex phenotypes that are reflected across different imaging channels. In such cases, optimal parameter combinations can be found using a simple machine learning model trained on a small ($n \cong 100$ cells per population) set of manually classified images (85).

An important decision during pooled cell-sorting screens is how to bin the phenotype of interest to maximize sensitivity to detect perturbation effects and exclude confounding factors. In most cases, two populations are isolated corresponding to low or high parameter values. Another strategy is to increase the number of bins on either side of the signal or to bin along the entire parameter. Finally, different parameters (such as the signal area and an image-derived parameter) can be combined for sorting, which is particularly useful when quantifying a protein's localization across cells with variable signal intensities. We recommend Reference 19 to readers trying to decide on the ideal binning strategy for the parameter of interest, irrespective of whether it is a traditional flow parameter or an image-derived parameter. Equally important are control samples: The gRNA distribution of the input gRNA plasmid library, the distribution of genome-integrated gRNA loci from cells before sorting, and the distribution of genome-integrated gRNA

loci from cells in the parent population of the actual phenotypic sorting gate are the available options. The frequently used analytical framework MAUDE requires one input sample and at least one sorted sample, plus the distribution of negative control gRNAs (19). The sizes of the sorting bins (percentage of cells in the sorted bin) must be known, and they should ideally contain equal cell numbers for statistical and practical reasons (e.g., easier sequencing sample preparation).

4.2.2. Library coverage. As is the case with all pooled genetic screens, sufficient coverage of the library is needed to ensure the sensitivity of the experiment. The relevant coverage in sorting-based screens is the number of cells after sorting, not the number of cells that were initially phenotyped. The lower the target population frequency, the more cells need to be phenotyped initially. Generally, 100× coverage after sorting is a good starting point for both FACS and ICS screens, as determined by comparing hit-calling performance at decreasing library coverage (85), or less systematically in many studies. Often, FACS-based screens are performed in two consecutive sorts with cell expansion between the two sorts (63, 71); however, such a repeated sorting strategy is time-consuming and might not be essential for many phenotypes of interest. Importantly, the nominal coverage of a FACS- or ICS-based screen will always be higher than the actual coverage since cell sorting is never 100% effective. Indeed, even in a perfectly calibrated cell sorter, only around 80% of the cells that were selected for isolation will actually be isolated, and this number will be lower if the instrument has not been set up properly (in a worst-case scenario, cells will be misclassified, causing false positives to be isolated).

4.2.3. Pilot experiments. Pilot experiments test the ability of the sorting strategy to separate negative and positive control perturbations and can be performed in an arrayed or pooled fashion. In addition, they help to estimate necessary coverage and determine whether the phenotype stays stable throughout the time needed for sorting (which often can be many hours). We recommend mixing cells from positive and negative control perturbations in different ratios, followed by sorting and genotyping of the isolated fractions. This allows comparison of the enrichment of the expected gRNA sequences over an input sample. In addition, the phenotypic purity of the isolated populations can be quantified using FACS (for FACS-based screens) or traditional microscopy (for ICS-based screens). Image-derived parameters are more variable compared to intensity-based parameters if recording the same cell multiple times because image-derived parameters can differ between measurements depending on the orientation of the cell relative to the optical path of the instrument. This is not the case for intensity-based parameters that make measurements from the entire cell, irrespective of the cell orientation. Therefore, the necessary coverage of ICS screens is equal to or higher than that of FACS screens, and the exact coverage needed depends on the phenotype of interest.

4.3. Perturbation Genotyping

Cell-sorting screens enrich cells with phenotypes of interest before perturbation genotyping. Perturbation genotyping is then done on isolated cell populations by targeted sequencing of genome-integrated gRNAs or associated barcodes. Genomic DNA is extracted from isolated cells, followed by PCR-based amplification of gRNA loci and sequencing. The abundance of gRNAs before or after sorting, between sorting fractions, and between the original DNA library and sorting bins can be used to identify gRNAs that were associated with the phenotype of interest (see Section 4.4).

4.4. Data Analysis

The analysis of cell sorting–based screens follows the same principles as those of standard growth-based fitness CRISPR screens. And ICS and FACS screens can be analyzed using the same tools

(**Figure 3**). Since phenotyping of cells happens during sorting, the only additional step after cell isolation is perturbation genotyping. Also, for ICS, no additional image analysis is required, as images are processed and analyzed on the fly during sorting (85). If the sorted population is compared against a single control, or two sorted bins are compared against each other, analysis of screen outputs is the same as for fitness screens or differential gene expression analysis, with many popular tools available (9, 16, 58, 61, 67). When multiple sorted populations are used, more nuanced methods become available. MAUDE infers a per-gRNA perturbation signal by building a generative model of the noisy stratification of cells into bins during sorting and estimating the most likely average fluorescence from the observed gRNA frequencies in the cell pools (19). The framework correctly captures the data-generating process, but its assumptions (e.g., of equal signal variance in each bin) should be updated as needed if modeling expertise is available. Regardless of the approach used, the last step of the analysis should be manual inspection of quality-control plots for the hits to avoid overinterpreting unexpected technical issues.

5. FUTURE TRENDS AND OUTLOOK

The technology for single-cell CRISPR screening, while still developing rapidly, is robust and broadly applied. How will the field evolve going forward?

5.1. More Biology with Complex Model Systems

While most pooled screening campaigns have been undertaken in cell lines, biological questions are best answered in the most meaningful cell types and model systems. Cells function in space, time, and local microenvironment, which cannot be modeled with simple in vitro models of isolated cell types. Recent in vivo screens with scRNA-seq readout in utero in mice (51), in allograft mouse models (44), and in brain organoids (30) demonstrated the compatibility of complex models with single-cell pooled screening and provided important insights into technical considerations for future studies. Similarly, cocultivation systems will allow the characterization of gene and cell function in defined contexts. Screens in stem cell–derived models, such as iPSC-derived cell lines, will resolve the temporal activity of the genome, from stem cell state to immature and fully mature cells.

5.2. Expanding the Perturbation Toolbox

Model organism studies have demonstrated how complex perturbations are a rich vein of novel insights. The complexity can take the shape of combinatorial perturbations of multiple genes in the same cell to identify epistatic interactions between genes and redundancy in gene regulation or drug–gene interactions to understand the mode of action and map modifiers of effect (**Figure 4**). In addition, precision editing and saturation editing will make it possible to test the functionality of each single nucleotide in a genome by installing defined genetic variants that saturate the genotypic space (**Figure 4**). Homology-directed repair CRISPR methods are still inefficient, resulting in only a small fraction of cells receiving an edit. To become compatible with single-cell pooled screening, these methods need to achieve much higher editing rates, for example, using specialized cell lines, such as repair-deficient cells or those expressing the CRISPR components at very high levels (30). Alternatively, methods need to be established that read out the gRNA, perturbation genotype at the edited locus, and the perturbation effect. These methods will allow more accurate perturbation genotyping and the exclusion of cells without successful edits.

In this review, we did not cover other perturbation systems beyond CRISPR, which are equally compatible with pooled single-cell screening. These include massively parallel reporter assays (MPRAs), deep mutational scanning assays (DMSAs), and ectopic gene expression systems.

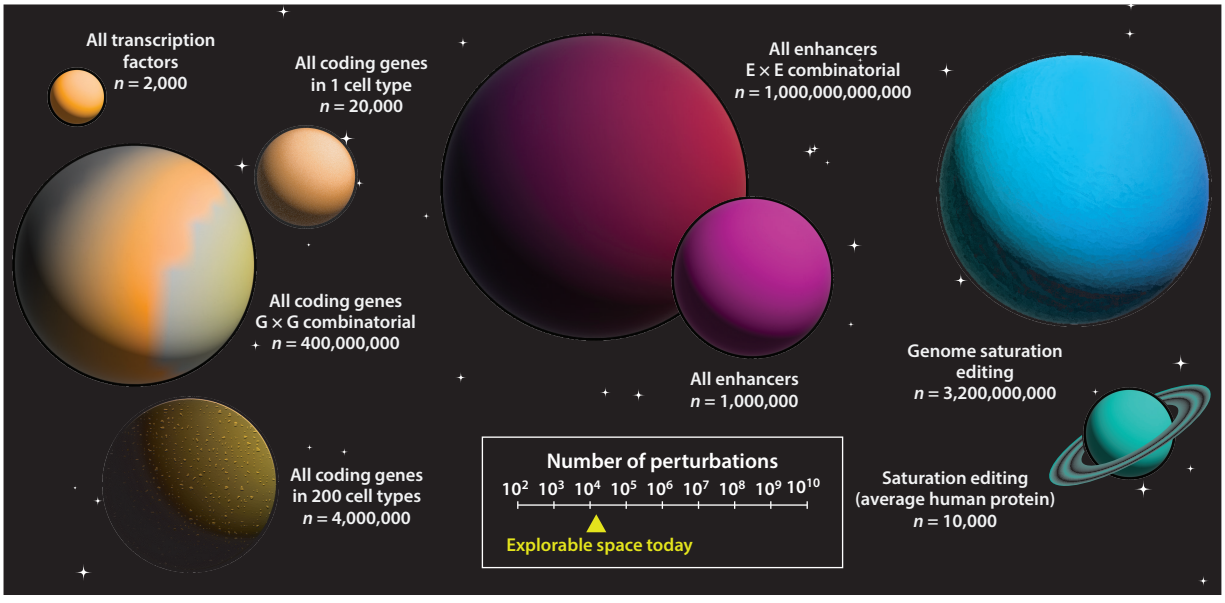

Figure 4

Differently scaled functional genomic screens illustrate the need for an additional two to three log scales in screening throughput. Planet diameters (in logarithmic scale) illustrate the complexity of different research questions in the functional genomic space. We set the explorable space (*yellow mark* in scale bar) to genome-wide screens targeting all protein-coding genes. Above this scale, we need new technological advancements with reduced experimental costs and intelligent experimental designs to prioritize gene sets for screens of the noncoding genome and combinatorial screens. As more and more data are collected, algorithms will become more accurate in predicting the effects of unseen perturbations across cell states. Saturation editing screens refer to the saturation of the amino acid (aa) coding space of an average-length human protein (472 aa protein × 20 aa), and saturation genome editing assumes that each nucleotide in the human genome is replaced once. Combinatorial perturbations are represented as all combinations of either two genes (G × G, 20,000 × 20,000 coding genes) or two enhancers (E × E, 1 million × 1 million enhancers). All transcription factors are mentioned as an example of a screen targeting all genes of a particular functional class.

MPRAs test thousands of regulatory elements on reporter gene expression by combining synthesized regulatory elements and reporter genes on a single expression construct. DMSAs generate variant libraries by means of gene synthesis followed by ectopic expression of those libraries within cells. Alternatively, DMSAs with precision genome or base editing allow the installation of genetic variants in an endogenous context. Finally, ectopic expression systems allow the overexpression of cDNA libraries or equivalent expression libraries. Altogether, these systems are compatible with pooled single-cell readouts by using genome-integrated barcodes that are linked to the expressed variant (102). Combinations of these systems, for example, overexpressing some genes while perturbing others, will be a near-inexhaustible hypothesis space to explore (70).

5.3. More Sophisticated Readouts

Single-cell technology is evolving to ultimately read out every omics layer from the same cell. Until this holy grail in the single-cell field is available, we will read out selected omics layers from single cells to uncover perturbation effects across these layers. For example, scRNA-seq with targeted readouts of the protein expression level has been applied for pooled CRISPR screening (31, 47, 69). High-throughput single-cell methods can also provide information about chromatin accessibility and genotypes at hundreds of loci, in addition to RNA-seq that captures transcriptional effects (45, 69, 78, 91). In particular, single-cell genotyping tools that sequence

transcriptomes and identify the effect of a CRISPR perturbation on DNA will be powerful in combination with precision genome editing tools (117). Finally, combined readouts of single-cell imaging and scRNA-seq will provide insights into the mode of action and phenotypic consequence of a perturbation from the same cell (86). The combination of pooled CRISPR screens in vivo with spatial transcriptomics readouts is another largely unexplored field that promises to identify the effects of genetic perturbations on tissue function and organization.

5.4. More Sophisticated Analysis Tools

Pooled single-cell screening analysis tool development is arguably outpacing the assays, yet much remains to be done. Many software options exist for all primary data analysis tasks, so alongside updating these two newest types of data modalities, we expect the focus in the near future to shift to data integration methodology and resources. There is an implementation tension between a precise but fragile model of the data generation process and a blunt but robust average change calculation that makes embedding many studies in a single aligned coordinate system tricky to optimize. Like the Human Cell Atlases, however, the integrated resources are very valuable to the community, so with multiple databases already emerging, we hope some will grow through the early problems to obtain long-term support and serve the needs of the field (73, 96).

5.5. Outlook

Five decades ago, major mechanisms of gene expression were still being figured out, and the first evidence of sequence-specific cutting emerged. A decade ago, single-cell experiments usually involved frog eggs and multiwell plates, while CRISPR had only just been proven to work in mammalian cells. Progress and cost reductions move fast, so we expect that perturbation effects of all single genes in static snapshots of homogenous cells will be gathered for most of the interesting cell types in the coming years. The combinatorial space of contexts and perturbations, as well as the function of regulatory elements in the noncoding genome, however, will probably never be densely measured due to the sheer number of hypotheses to be tested (**Figure 4**). Most fruitful outcomes will stem from the judicious choice to add to studies the dimension of time through differentiation and development, multiple perturbations to elucidate causal links and uncover pathway structures, and controlled cell type compositions to systematically measure the impact of neighboring environments.

DISCLOSURE STATEMENT

L.M.S. is co-founder and shareholder of Sophia Genetics, LevitasBio, and Recombia Biosciences. Other than that, the authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Szen Toh, Claudia Feng, and Juliane Weller for comments on the manuscript. L.P. was funded by the Wellcome Sanger Institute (108413/A/15/D). D.S. was supported by a fellowship from the EMBL Interdisciplinary Postdoctoral (EIPOD) program (Marie Skłodowska-Curie Actions COFUND grant agreement 664726). L.M.S. was supported by grants from the European Research Council (Advanced Investigator Grants AdG-294542 and AdG-742804).

LITERATURE CITED

1. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, et al. 2016. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167:1867–82.e21

2. Adamson B, Norman TM, Jost M, Weissman JS. 2018. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. bioRxiv 298349. **https://doi.org/10.1101/298349**
3. Alerasool N, Segal D, Lee H, Taipale M. 2020. An efficient KRAB domain for CRISPRi applications in human cells. *Nat. Methods* 17:1093–96
4. Andrews TS, Kiselev VY, McCarthy D, Hemberg M. 2021. Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nat. Protoc.* 16:1–9
5. Anzalone AV, Koblan LW, Liu DR. 2020. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38:824–44
6. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, et al. 2019. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576:149–57
7. Behan FM, Iorio F, Picco G, Goncalves E, Beaver CM, et al. 2019. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* 568:511–16
8. Bock C, Datlinger P, Chardon F, Coelho MA, Dong MTB, et al. 2022. High-content CRISPR screening. *Nat. Rev. Methods Primers* 2:8
9. Bodapati S, Daley TP, Lin X, Zou J, Qi LS. 2020. A benchmark of algorithms for the analysis of pooled CRISPR screens. *Genome Biol.* 21:62
10. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357:661–67
11. Cetin R, Wegner M, Luwisch L, Saud S, Achmedov T, et al. 2023. Optimized metrics for orthogonal combinatorial CRISPR screens. *Sci. Rep.* 13:7405
12. Chen PJ, Liu DR. 2023. Prime editing for precise and highly versatile genome manipulation. *Nat. Rev. Genet.* 24:161–77
13. Cheng J, Lin G, Wang T, Wang Y, Guo W, et al. 2023. Massively parallel CRISPR-based genetic perturbation screening at single-cell resolution. *Adv. Sci.* 10:e2204484
14. Cheng L, Li YC, Qi Q, Xu P, Feng RP, et al. 2021. Single-nucleotide-level mapping of DNA regulatory elements that control fetal hemoglobin expression. *Nat. Genet.* 53:869–80
15. Cleary B, Cong L, Cheung A, Lander ES, Regev A. 2017. Efficient generation of transcriptomic profiles by random composite measurements. *Cell* 171:1424–36.e18
16. Colic M, Hart T. 2021. Common computational tools for analyzing CRISPR screens. *Emerg. Top. Life Sci.* 5:779–88
17. Cossarizza A, Chang H-D, Radbruch A, Acs A, Adam D, et al. 2019. Guidelines for the use of flow cytometry and cell sorting in immunological studies (second edition). *Eur. J. Immunol.* 49:1457–973
18. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, et al. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14:297–301
19. de Boer CG, Ray JP, Hacohen N, Regev A. 2020. MAUDE: inferring expression changes in sorting-based CRISPR screens. *Genome Biol.* 21:134
20. Dede M, McLaughlin M, Kim E, Hart T. 2020. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol.* 21:262
21. Deshpande R, Nelson J, Simpkins SW, Costanzo M, Piotrowski JS, et al. 2017. Efficient strategies for screening large-scale genetic interaction networks. bioRxiv 159632. **https://doi.org/10.1101/159632**
22. Dixit A, Parnas O, Li B, Chen J, Fulco CP, et al. 2016. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167:1853–66.e17
23. Doench JG. 2018. Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* 19:67–80
24. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnol.* 34:184–91
25. Drager NM, Sattler SM, Huang CT, Teter OM, Leng K, et al. 2022. A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states. *Nat. Neurosci.* 25:1149–62
26. Erwood S, Bily TMI, Lequyer J, Yan J, Gulati N, et al. 2022. Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* 40:885–95
27. Feldman D, Singh A, Garrity AJ, Blainey PC. 2018. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. bioRxiv 262121. **https://doi.org/10.1101/262121**

28. Feldman D, Singh A, Schmid-Burgk JL, Carlson RJ, Mezger A, et al. 2019. Optical pooled screens in human cells. *Cell* 179:787–99.e17
29. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. 2014. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* 513:120–23
30. Fleck JS, Jansen SMJ, Wollny D, Zenk F, Seimiya M, et al. 2022. Inferring and perturbing cell fate regulomes in human brain organoids. *Nature*. **https://doi.org/10.1038/s41586-022-05279-8**
31. Frangieh CJ, Melms JC, Thakore PI, Geiger-Schuller KR, Ho P, et al. 2021. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. *Nat. Genet.* 53:332–41
32. Funk L, Su KC, Ly J, Feldman D, Singh A, et al. 2022. The phenotypic landscape of essential human genes. *Cell* 185:4634–53.e22
33. Gasperini M, Hill AJ, McFaline-Figueroa JL, Martin B, Kim S, et al. 2019. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* 176:377–90
34. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, et al. 2022. A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* 40:163–66
35. Geiger-Schuller K, Eraslan B, Kuksenko O, Dey KK, Jagadeesh KA, et al. 2023. Systematically characterizing the roles of E3-ligase family members in inflammatory responses with massively parallel Perturb-seq. bioRxiv 2023.01.23.525198. **https://doi.org/10.1101/2023.01.23.525198**
36. Gier RA, Budinich KA, Evitt NH, Cao Z, Freilich ES, et al. 2020. High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nat. Commun.* 11:3455
37. Guna A, Page KR, Replogle JR, Esantsi TK, Wang ML, et al. 2023. A dual sgRNA library design to probe genetic modifiers using genome-wide CRISPRi screens. bioRxiv 2023.01.22.525086. **https://doi.org/10.1101/2023.01.22.525086**
38. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, et al. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184:3573–87.e29
39. Hasle N, Cooke A, Srivatsan S, Huang H, Stephany JJ, et al. 2020. High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* 16:e9442
40. Heumos L, Schaar AC, Lance C, Litinetskaya A, Drost F, et al. 2023. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* 24:550–72
41. Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, et al. 2018. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* 15:271–74
42. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, et al. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33:510–17
43. Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, et al. 2016. Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* 5:e19760
44. Hou J, Liang S, Xu C, Wei Y, Wang Y, et al. 2022. Single-cell CRISPR immune screens reveal immunological roles of tumor intrinsic factors. *NAR Cancer* 4:zcac038
45. Hou Y, Guo H, Cao C, Li X, Hu B, et al. 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* 26:304–19
46. Huang TP, Newby GA, Liu DR. 2021. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat. Protoc.* 16:1089–128
47. Hwang B, Lee DS, Tamaki W, Sun Y, Ogorodnikov A, et al. 2021. SCITO-seq: single-cell combinatorial indexed cytometry sequencing. *Nat. Methods* 18:903–11
48. Ibraheim R, Tai PWL, Mir A, Javeed N, Wang J, et al. 2021. Self-inactivating, all-in-one AAV vectors for precision Cas9 genome editing via homology-directed repair in vivo. *Nat. Commun.* 12:6267
49. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, et al. 2014. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–79
50. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, et al. 2016. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167:1883–96.e15
51. Jin X, Simmons SK, Guo A, Shetty AS, Ko M, et al. 2020. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* 370:eaaz6063

52. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. 2023. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* 614:742–51
53. Kampmann M. 2018. CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem. Biol.* 13:406–16
54. Kanfer G, Sarraf SA, Maman Y, Baldwin H, Dominguez-Martin E, et al. 2021. Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J. Cell Biol.* 220:e202006180
55. Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, et al. 2015. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* 12:401–3
56. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, et al. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–201
57. Koeppel J, Peets EM, Weller J, Pallaseni A, Liberante F, Parts L. 2021. Predicting efficiency of writing short sequences into the genome using prime editing. bioRxiv 2021.11.10.468024. **https://doi.org/10.1101/2021.11.10.468024**
58. Li W, Xu H, Xiao T, Cong L, Love MI, et al. 2014. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15:554
59. Liscovitch-Brauer N, Montalbano A, Deng J, Mendez-Mancilla A, Wessels HH, et al. 2021. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* 39:1270–77
60. Lopez R, Tagasovska N, Ra S, Cho K, Pritchard JK, Regev A. 2022. Learning causal representations of single cells via sparse mechanism shift modeling. arXiv:2211.03553 [q-bio.GN]
61. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550
62. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–14
63. Mair B, Aldridge PM, Atwal RS, Philpott D, Zhang M, et al. 2019. High-throughput genome-wide phenotypic screening via immunomagnetic cell sorting. *Nat. Biomed. Eng.* 3:796–805
64. Mandegar MA, Huebsch N, Frolov EB, Shin E, Truong A, et al. 2016. CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* 18:541–53
65. Marshall JL, Doughty BR, Subramanian V, Guckelberger P, Wang Q, et al. 2020. HyPR-seq: single-cell quantification of chosen RNAs via hybridization and sequencing of DNA probes. *PNAS* 117:33404–13
66. McCarty NS, Graham AE, Studena L, Ledesma-Amaro R. 2020. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* 11:1281
67. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, et al. 2017. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* 49:1779–84
68. Michlits G, Hubmann M, Wu SH, Vainorius G, Budusan E, et al. 2017. CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat. Methods* 14:1191–97
69. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, et al. 2019. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* 16:409–12
70. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, et al. 2018. Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* 36:179–89
71. Panganiban RA, Park HR, Sun M, Shumyatcher M, Himes BE, Lu Q. 2019. Genome-wide CRISPR screen identifies suppressors of endoplasmic reticulum stress-induced apoptosis. *PNAS* 116:13384–93
72. Peets EM, Crepaldi L, Zhou Y, Allen F, Elmentaite R, et al. 2019. Minimized double guide RNA libraries enable scale-limited CRISPR/Cas9 screens. bioRxiv 859652. **https://doi.org/10.1101/859652**
73. Peidli S, Green TD, Shen C, Gross T, Min J, et al. 2023. scPerturb: harmonized single-cell perturbation data. bioRxiv 2022.08.20.504663. **https://www.biorxiv.org/content/10.1101/2022.08.20.504663v3**
74. Perez AR, Sala L, Perez RK, Vidigal JA. 2021. CSC software corrects off-target mediated gRNA depletion in CRISPR-Cas9 essentiality screens. *Nat. Commun.* 12:6461
75. Pierce SE, Granja JM, Greenleaf WJ. 2021. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat. Commun.* 12:2969
76. Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, et al. 2020. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* 38:954–61

77. Replogle JM, Saunders RA, Pogson AN, Hussmann JA, Lenail A, et al. 2022. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* 185:2559–75.e28
78. Rodriguez-Meira A, Buck G, Clark SA, Povinelli BJ, Alcolea V, et al. 2019. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol. Cell* 73:1292–305.e8
79. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* 360:176–82
80. Roth TL, Li PJ, Blaeschke F, Nies JF, Apathy R, et al. 2020. Pooled knockin targeting for genome engineering of cellular immunotherapies. *Cell* 181:728–44.e21
81. Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, et al. 2019. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* 176:361–76.e17
82. Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, et al. 2018. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* 9:5416
83. Schmidt R, Steinhart Z, Layeghi M, Freimer JW, Bueno R, et al. 2022. CRISPR activation and interference screens decode stimulation responses in primary human T cells. *Science* 375:eabj4008
84. Schraivogel D, Gschwind AR, Milbank JH, Leonce DR, Jakob P, et al. 2020. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* 17:629–35
85. Schraivogel D, Kuhn TM, Rauscher B, Rodríguez-Martínez M, Paulsen M, et al. 2022. High-speed fluorescence image–enabled cell sorting. *Science* 375:315–20
86. Schraivogel D, Steinmetz LM. 2023. Cell sorters see things more clearly now. *Mol. Syst. Biol.* 19:e11254
87. Schumann K, Raju SS, Lauber M, Kolb S, Shifrut E, et al. 2020. Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. *Nat. Immunol.* 21:1456–66
88. Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, et al. 2018. Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. *Cell* 175:1958–71.e15
89. Smits AH, Ziebell F, Joberty G, Zinn N, Mueller WF, et al. 2019. Biological plasticity rescues target activity in CRISPR knock outs. *Nat. Methods* 16:1087–93
90. Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao JY, et al. 2020. Massively multiplex chemical transcriptomics at single-cell resolution. *Science* 367:45–51
91. Stoeckius M. 2017. Large-scale simultaneous measurement of epitopes and transcriptomes in single cells. *Nat. Methods* 14:865–68
92. Stuart T, Satija R. 2019. Integrative single-cell analysis. *Nat. Rev. Genet.* 20:257–72
93. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171:1437–52.e17
94. Sun N, Petiwala S, Wang R, Lu C, Hu M, et al. 2019. Development of drug-inducible CRISPR-Cas9 systems for large-scale functional screening. *BMC Genom.* 20:225
95. Svensson V, Vento-Tormo R, Teichmann SA. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13:599–604
96. Tian R, Abarientos A, Hong J, Hashemi SH, Yan R, et al. 2021. Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. *Nat. Neurosci.* 24:1020–34
97. Tian R, Gachechiladze MA, Ludwig CH, Laurie MT, Hong JY, et al. 2019. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* 104:239–55.e12
98. Ting PY, Parker AE, Lee JS, Trussell C, Sharif O, et al. 2018. Guide Swap enables genome-scale pooled CRISPR-Cas9 screening in human primary cells. *Nat. Methods* 15:941–46
99. Tran V, Papalexi E, Schroeder S, Kim G, Sapre A, et al. 2022. High sensitivity single cell RNA sequencing with split pool barcoding. bioRxiv 2022.08.27.505512. **https://doi.org/10.1101/2022.08.27.505512**
100. Truong DJ, Kuhner K, Kuhn R, Werfel S, Engelhardt S, et al. 2015. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* 43:6450–58
101. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, et al. 2017. Defining a cancer dependency map. *Cell* 170:564–76.e16
102. Ursu O, Neal JT, Shea E, Thakore PI, Jerby-Arnon L, et al. 2022. Massively parallel phenotyping of coding variants in cancer with Perturb-seq. *Nat. Biotechnol.* 40:896–905

103. Usluer S, Hallast P, Crepaldi L, Zhou Y, Urgo K, et al. 2022. Optimised whole-genome CRISPR interference screens identify ARID1A-dependent growth regulators in human induced pluripotent stem cells. *Stem Cell Rep*. 18(5):1061–74
104. Vinceti A, Perron U, Trastulla L, Iorio F. 2022. Reduced gene templates for supervised analysis of scale-limited CRISPR-Cas9 fitness screens. *Cell Rep*. 40:111145
105. Walton RT, Singh A, Blainey PC. 2022. Pooled genetic screens with image-based profiling. *Mol. Syst. Biol.* 18:e10768
106. Wang C, Lu T, Emanuel G, Babcock HP, Zhuang X. 2019. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *PNAS* 116:10842–851
107. Wessels HH, Mendez-Mancilla A, Hao Y, Papalexi E, Mauck WM 3rd, et al. 2023. Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq. *Nat. Methods* 20:86–94
108. Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*. 19:15
109. Wroblewska A, Dhainaut M, Ben-Zvi B, Rose SA, Park ES, et al. 2018. Protein barcodes enable high-dimensional single-cell CRISPR screens. *Cell* 175:1141–55.e16
110. Xie S, Duan J, Li B, Zhou P, Hon GC. 2017. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell* 66:285–99.e5
111. Xu X, Chemparathy A, Zeng L, Kempton HR, Shang S, et al. 2021. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* 81:4333–45.e4
112. Yamano S, Dai J, Moursi AM. 2010. Comparison of transfection efficiency of nonviral gene transfer reagents. *Mol. Biotechnol.* 46:287–300
113. Yan X, Stuurman N, Ribeiro SA, Tanenbaum ME, Horlbeck MA, et al. 2021. High-content imaging-based pooled CRISPR screens in mammalian cells. *J. Cell Biol.* 220:e202008158
114. Yang L, Zhu Y, Yu H, Cheng X, Chen S, et al. 2020. scMAGeCK links genotypes with multiple phenotypes in single-cell CRISPR screens. *Genome Biol*. 21:19
115. Yao D, Binan L, Bezney J, Simonton B, Freedman J, et al. 2023. Compressed Perturb-seq: highly efficient screens for regulatory circuits using random composite perturbations. bioRxiv 2023.01.23.525200. **https://doi.org/10.1101/2023.01.23.525200**
116. Yeo NC, Chavez A, Lance-Byrne A, Chan Y, Menn D, et al. 2018. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* 15:611–16
117. Yu L, Wang X, Mu Q, Tam SST, Loi DSC, et al. 2023. scONE-seq: A single-cell multi-omics method enables simultaneous dissection of phenotype and genotype heterogeneity from frozen tumors. *Sci. Adv.* 9:eabp8901
118. Zappia L, Theis FJ. 2021. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol*. 22:301
119. Zetsche B, Volz SE, Zhang F. 2015. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* 33:139–42
120. Zhang H, Li T, Sun Y, Yang H. 2021. Perfecting targeting in CRISPR. *Annu. Rev. Genet.* 55:453–77
121. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, et al. 2017. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049